

ABSTRAK

Berdasarkan data dari Pusat Data dan Informasi Kementerian Kesehatan RI, di tahun 2012 sekitar 8,2 juta kasus kematian disebabkan oleh kanker. Perkembangan terakhir menunjukkan bahwa teknologi DNA *microarray* mampu menangani masalah deteksi kanker sejak dini, namun kelemahan utama dari *microarray* adalah masalah *curse of dimensionality*.

Analysis of Variance (ANOVA) merupakan salah satu metode seleksi fitur yang dapat mengatasi kelemahan *microarray*. ANOVA dapat menemukan pasangan gen informatif yang dapat membantu dalam proses pengklasifikasian yang dilakukan oleh Support Vector Machine (SVM). Dalam SVM, kernel *trick* saat *learning* model sangat membantu dalam mengatasi masalah *feature space*. Pemilihan kernel berpengaruh terhadap akurasi yang dihasilkan. Melalui serangkaian proses seperti perhitungan korelasi, seleksi fitur dan pengklasifikasian menggunakan SVM, didapatkan akurasi dari empat *dataset* yang digunakan. Untuk *dataset* leukimia dan *ovarian cancer*, akurasi terbesar dihasilkan oleh kernel polynomial yaitu sebesar 100% dan 97,54%. Sedangkan untuk *dataset lung cancer* akurasi terbesar diperoleh dari kernel linear yaitu sebesar 100% dan untuk *dataset colon tumor* akurasi terbesar diperoleh dari kernel RBF sebesar 85,15%. Perbedaan kernel yang menghasilkan akurasi tertinggi pada setiap *dataset* sangat bergantung kepada karakteristik *dataset* kanker itu sendiri.

Kata kunci: deteksi kanker, DNA *microarray*, reduksi dimensi, korelasi, analysis of variance, support vector machine, *kernel trick*